

# 新零售无人智能售货机商务数据分析报告

## 目录

- 一、 项目背景.....1
  - 1. 行业背景.....1
  - 2. 项目目标.....1
- 二、 数据说明.....2
  - 1. 数据来源.....2
  - 2. 数据采集.....2
  - 3. 数据量.....2
- 三、 数据预处理与分析.....2
  - 1. 数据表检查.....2
  - 2. 数据预处理.....2
  - 3. 数据表提取.....3
  - 4. 数据计算.....4
- 四、 数据可视化分析.....6
- 五、 自动售货机画像.....14
  - 1. 贴标签.....14
  - 2. 生成各自动售货机画像.....15
- 六、 销售额预测.....19
  - 1. 预测未来销售额的原理.....19
  - 2. 关于本案例中销售额的预测.....23
- 七、 项目总结.....24

## 一、 项目背景

### 1. 行业背景

2016 年 10 月的阿里巴巴云栖大会上，马云第一次提出了新零售的概念，并预言在未来十年电子商务将死，新零售将成为新的引领经济发展的强力商业模式。

新零售是以互联网技术为基础，以大数据为导向，充分进行供应链整合，线上线下综合销售的商业模式，具体又可以分为无人售货、综合门店、便捷连锁、无人货架等等不同模式。

新零售无人智能售货机以更贴近客户为宗旨，在各类型场景满足客户需求，从销售端带动生产端，提高产品附加值，并通过大数据、先进科技、专业设备及严格的流程把控，整合供应链管，降低成本和风险，从而在不做超出市场定价的前提下实现较高毛利。

自动售货机内商品的供给频率、种类选择、供给量站点选择等是自动售货机运营者需要重点关注的问题。因此，科学的商业数据分析能够帮助经营者了解用户需求，握商品需求量，为用户提供精准贴心的服务，是握经营方向的重要手段，对自动售货机这一营销模式的发展有非常重要的意义。

### 2. 项目目标

- a) 根据自动售货机的经营特点，对经营指标数据、商品营销数据及市场需求进行分析，完成对销量、库存、盈利三个方面各项指标的计算，按要求绘制对应图表，实现数据可视化；
- b) 为每台售货机所销售的商品贴上标签，使其能够更好地展现销售

商品的特征；

- c) 根据数据以及计算，预测每台售货机的每大类商品在 2018 年 1 月的交易额。

## 二、 数据说明

### 1. 数据来源

数据来源：某商场中不同地点 A、B、C、D、E 处安置的自动售货机

数据时间：2017 年 1 月 1 日至 2017 年 12 月 31 日

### 2. 数据采集

商品信息采集：一共有 315 种商品，并已将其分为饮料类与非饮料类。

### 3. 数据量

销售记录：70680 条记录

## 三、 数据预处理与分析

### 1. 数据表检查

在数据预处理之前，先了解了数据表的形状以及数据表的数据类型，得知附件一的数据一共有 70680 行 9 列，并且有一列需要转换格式，这为接下来的数据预处理做好准备。

### 2. 数据预处理

数据预处理常用的方法有数据清洗、数据合并、数据转换等，预处理步骤不分先后，可以反复进行。本次分析报告的数据预处理主要集中在清洗数据上，包括检查与处理缺失值、检查与处理重复值、检

查与处理异常值。

## 2.1 检查与处理缺失值

检查附件一数据的完整性，查找数据的缺失值，发现数据中有一条记录的支付时间缺失。对于确缺失的处理，有两种方式，第一是直接删除缺失值所在的行或者列，第二则是用其他数据的平均值，众数，或者上下值来填充这个数据表格中的缺失值。对于本项目的缺失值，适用于第一种方式，利用 `dropna()` 函数直接将缺失值所在的行。

## 2.2 检查与处理重复值

对于重复值，本项目主要考虑记录重复，可以选择直接删除来提高预测的精准度，不过该数据表没有重复值，因此不用删除数据。

## 2.3 检查与处理异常值

对于异常值，其处理要视情况而定，因为有些异常值可能蕴含着有用的信息，没用的异常值可以进行删除，经过检查发现，该数据表中有一条记录的支付时间异常，2017 年没有 2 月 29 日，该异常值没有用处，删除该条记录。

除此了清洗数据外，本次项目报告也涉及了合并数据、转换数据。比如，在数据可视化分析之前，将附件 2 商品的分类与清洗后对数据合并在同一张表格里，还有更改数据格式，将“支付时间”的类型从 `object` 转化为 `datetimes`。

## 3. 数据表提取

在数据预处理后，通过 `iloc()` 函数提取了每台售货机对应的

销售数据，保存在 CSV 文件中，文件名分别为“task1-1A.csv”、“task1-1B.csv”、“task1-1C.csv”、“task1-1D.csv”、“task1-1E.csv”。

#### 4. 数据计算

##### 4.1 每台售货机 2017 年 5 月份对交易额、订单量以及所有售货机交易总额和订单总量

	交易额	订单量
A	2816.8	756
B	3223.8	869
C	2993.8	789
D	2047.9	564
E	4996.1	1292
合计	16078.4	4270

表 1 每台售货机 2017 年 5 月份对交易额、订单量

##### 4.2 每台售货机每月的每单平均交易额与日均订单量

每台售货机每月的每单平均交易额					
地点 月份	A	B	C	D	E
1	3.600597	3.416667	3.646966	3.427799	3.773729
2	3.280702	3.139459	3.567633	2.975177	3.404264
3	3.177255	3.346415	3.526616	3.479167	4.134286
4	3.620134	3.502156	3.754087	3.529571	3.586927

5	3. 725926	3. 709781	3. 794423	3. 631028	3. 866950
6	3. 505333	3. 550108	3. 580553	3. 631442	3. 489780
7	3. 576261	3. 617971	3. 533246	3. 497792	3. 493358
8	3. 120270	3. 399796	3. 621922	3. 179720	3. 612960
9	3. 664423	3. 489341	3. 641776	3. 454934	3. 668191
10	3. 487987	3. 599062	3. 692554	3. 476560	3. 506986
11	3. 737759	3. 493255	3. 646629	3. 463306	3. 771972
12	3. 398952	3. 386697	3. 566877	3. 306494	3. 703690

表 2 每台售货机每月的每单平均交易额

每台售货机每月的日均订单量					
地 点 月份	A	B	C	D	E
1	10. 806452	11. 806452	12. 225806	8. 354839	11. 419355
2	4. 071429	6. 607143	7. 392857	5. 035714	9. 214286
3	8. 225806	8. 548387	8. 483871	6. 193548	11. 290323
4	14. 900000	20. 100000	24. 466667	14. 766667	29. 833333
5	24. 387097	28. 032258	25. 451613	18. 193548	41. 677419
6	55. 633333	61. 866667	62. 733333	34. 666667	86. 433333
7	15. 354839	11. 129032	24. 645161	10. 225806	26. 225806
8	21. 483871	31. 645161	40. 612903	23. 064516	57. 00000
9	34. 666667	58. 166667	55. 933333	32. 766667	137. 800000

10	50.483871	65.354839	71.483871	38.258065	89.580645
11	38.666667	67.700000	64.766667	40.333333	167.333333
12	64.612903	71.290323	76.741935	53.645161	104.903226

表 3 每台售货机每月的日均订单量

#### 四、 数据可视化分析

##### 1. 2017 年 6 月销量前五的商品及其销量柱状图

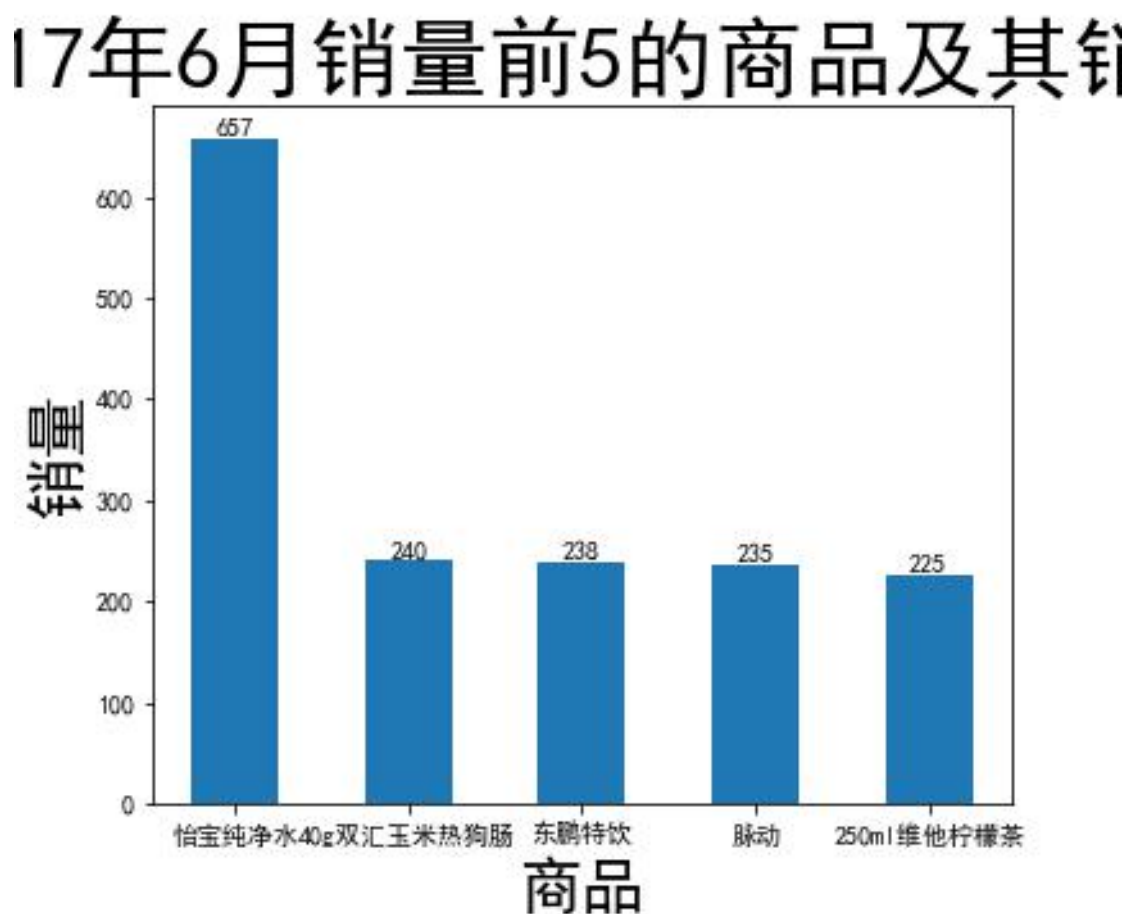


图 1 2017 年 6 月销量前五的商品及其销量柱状图

由图 1 可知，2017 年 6 月销量前五的商品分别为怡宝纯净水、40g 双汇玉米热狗肠、东鹏特饮、脉动、250ml 维他柠檬茶，其中四种商品是饮料类，可见饮料类是大多数人的需求，并且怡宝纯净水的销售量大约是所有其他商品销售量的 2.5 倍，其他四种商品的销售量不相

上下，这种情况应该与当时的天气有关，6 月份的时候天气比较炎热，人们容易口渴，所以人们比较倾向于购买饮料类，特别是解渴的饮料，就如怡宝纯净水，应增大怡宝纯净水库存量。

2. 各售货机每月总交易额折线图与月环比增长率柱状图

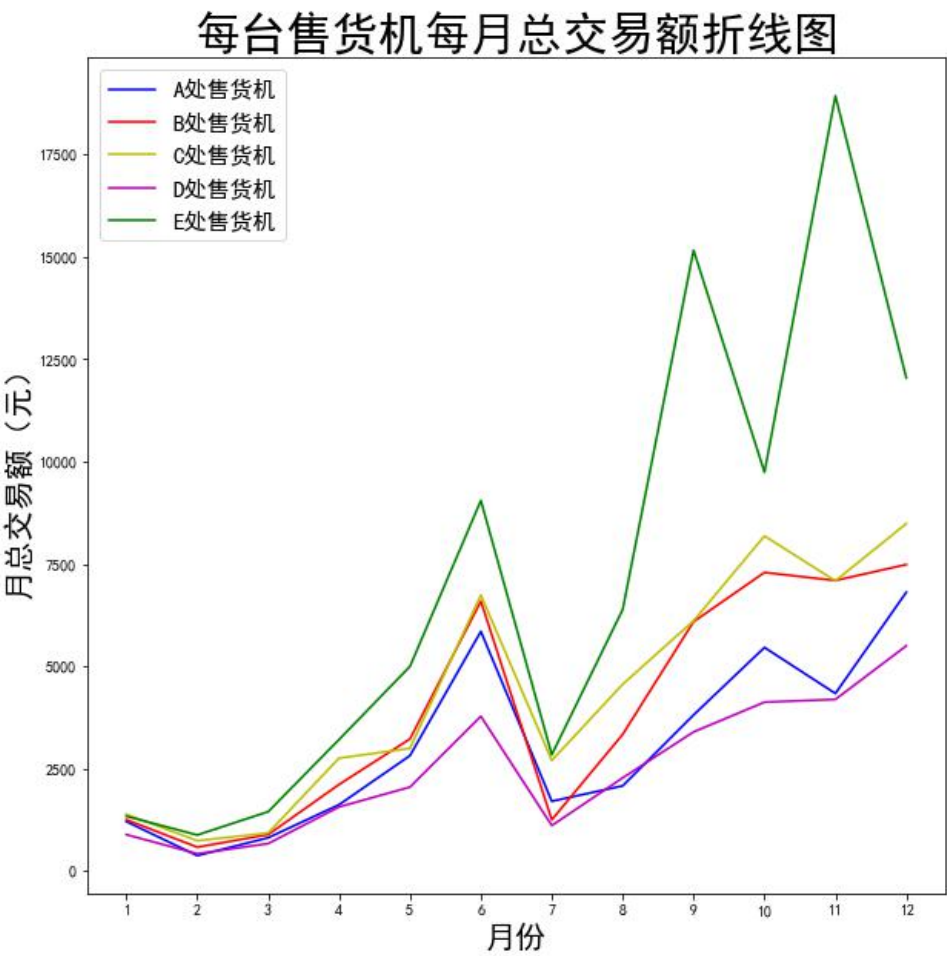


图 2 各售货机每月总交易额折线图



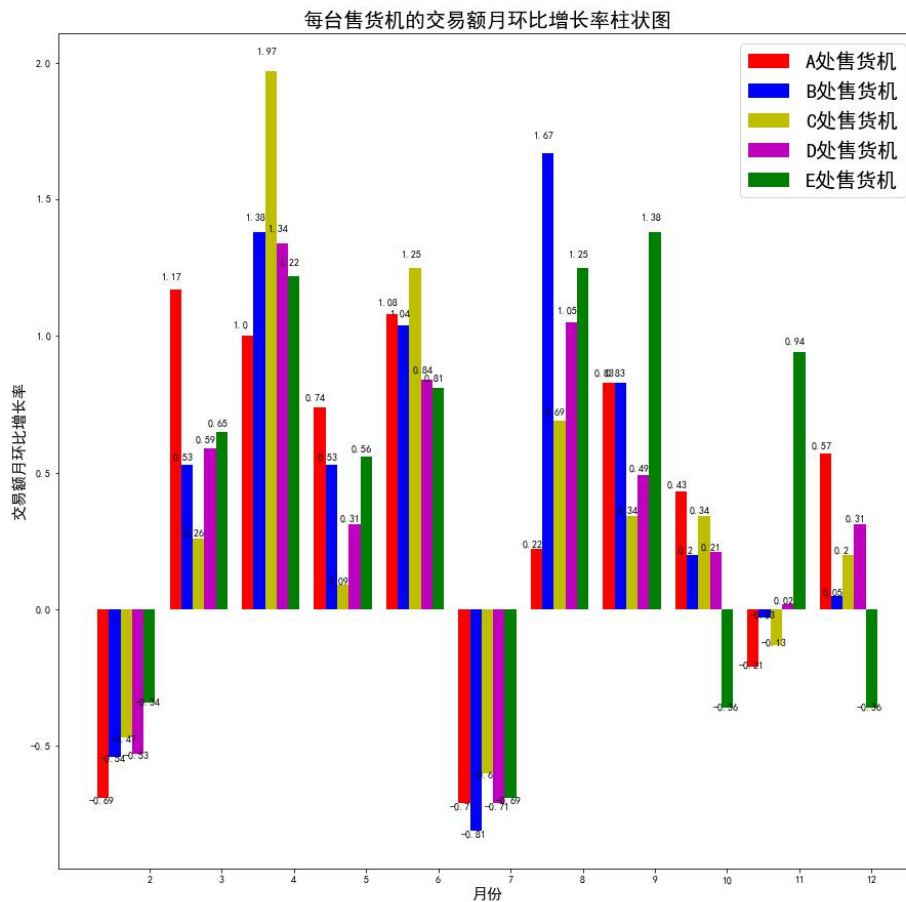


图 3 各售货机每月总交易额的月环比增长率柱状图

通过图 2，可以得知 2017 年各售货机每月总交易额的变化趋势，总体上各售货机每月总交易额的变化趋势基本一致，具有相同的转折点和极点。

2 月份是所有售货机每月总交易额最少的极点，2 月份正好是我国春节期间，由于我国春节的习俗，大家都会在春节前买许许多多的商品放在家里，并且人们都去探望亲朋好友，很少人去商场，因此自动售货机的订单量减少，也即交易额减少。过了 2 月后，交易总额慢慢回升，3、4、5、6 月份的交易额遂渐升高，但 7 月份的交易总额

剧减，这与当月天气有关，7 月份正是酷热时期，面对炎热，大多数人们选择在家享受空调，而减少外出的次数，那么光顾售货机的次数也会减少，因此导致 7 月份的交易总额剧减。过了 7 月后，8、9、10 月交易额逐渐增加，然后从 11 份开始交易总额有明显的波折，11 月的交易总额下降，然后 12 月的交易总额又上升，但是有个例外，E 处的售货机是 10 月份的交易总额下降，11 月的交易总额上升，并且上升速度非常快，达到 E 处售货机交易额的最高峰，甚至是所有售货机月交易额中的最大值，这可能是因为 11 月份时 E 处曾经举办了某大型活动，使 E 处的人流量巨增。

图 3 中，交易额月环比增长率是本月份的交易额与上一个月的交易额的差值，然后比上上一个月份的交易额的值，因此第一个月份没有环比增长率。通过每台售货机交易额月环比增长率的柱状图和上面标记的数据更能科学有力地比较每台售货机月交易额的情况，可知 2 月、7 月、11 月的交易额月环比增长率为负值，即它们的交易额均比上一个月份少。

### 3. 每台售货机毛利润占总毛利润的饼图

## 每台售货机毛利润占总毛利润的饼图



图 4 每台售货机毛利润占总毛利润的饼图

图 4 中，假设饮料类的毛利率为 25%，非饮料类 20%，每台售货机的毛利润由每台售货机的饮料类和非饮料类的总交易额分别乘以相应的毛利率再相加得出，由此画出每台售货机毛利润占总毛利润的饼图。由饼图可知，A 处售货机所赚的毛利润占比为 14.6%，B 处售货机所赚的毛利润占比为 18.8%，C 处售货机所赚的毛利润占比为 21.5%，D 处售货机所赚的毛利润占比为 11.6%，E 处售货机所赚的毛利润占比为 33.6%，对比得：E 处售货机所赚的毛利润占比最高，接着依次递减的是 C 处售货机、B 处售货机、A 处售货机，D 处售货机

的毛利润占比最小。可猜测 E 处是商场中人流较大的地区，因此 E 处自动售货机在此处的商机较大，应增加 E 处自动售货机的库存，及时供给，添加商品种类；相比之下，D 处是商场中人流较小的地区，对于 D 处售货机，可以重新考虑站点的选择。

4. 每月交易总额均值的气泡图

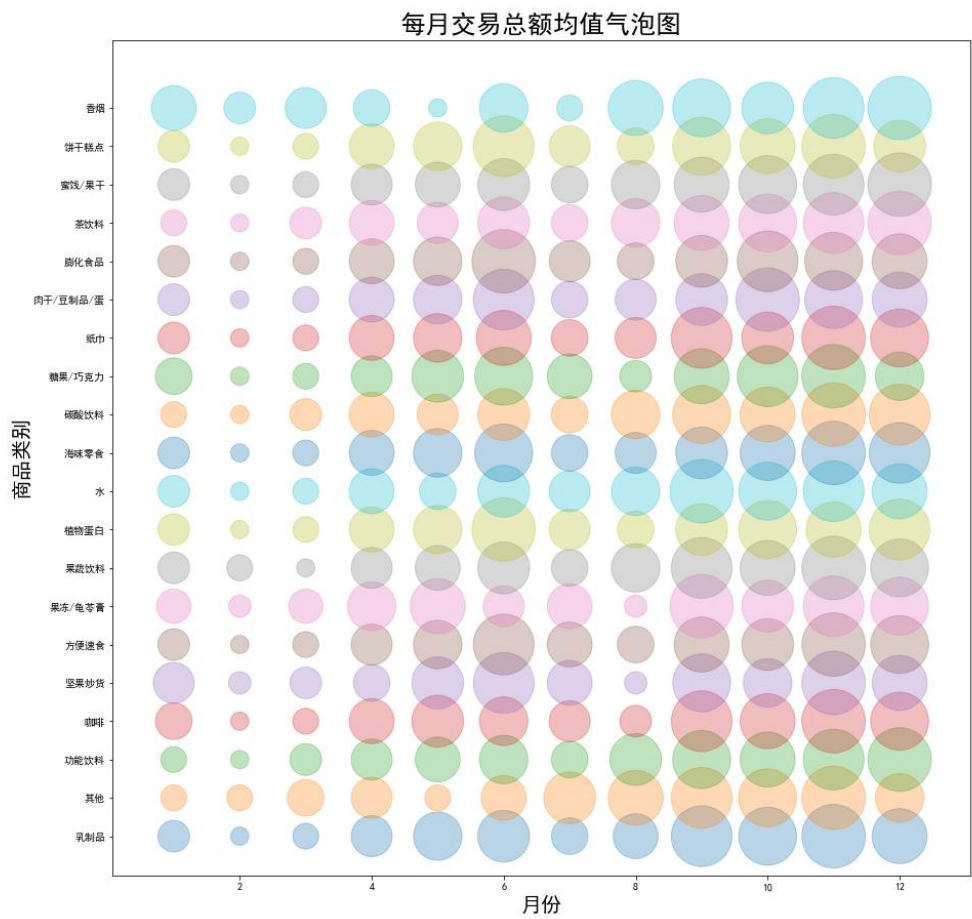


图 5 每月交易总额均值的气泡图

图 5 是每月交易总额均值的气泡图，横轴是月份，纵轴是商品的二级类目，气泡的大小对应交易额均值的多少，即可以通过比较气泡

的大小来表示交易额均值的大小。先纵向观察气泡的大小，可发现 2 月上的气泡最小，即 2 月的交易总额均值整体最小，而 6、9、10、11、12 月份的气泡较大，即 6、9、10、11、12 月份的交易总额均值整体都很大；然后横向观察可发现，各类商品的月交易总额均值较大的月份主要集中在 6、9、10、11、12 月，其他月份的交易总额均值较小。此外，各类商品的月交易总额均值的最大值和最小值相近，没有出现哪种商品的月交易额均值异常大或者异常小。结合图 2，还能发现各类商品的月交易总额均值的变化趋势与各台售货机每月交易总额折线图相似。

5. 售货机 C 在 2017 年 6、7、8 月份订单量的热力图

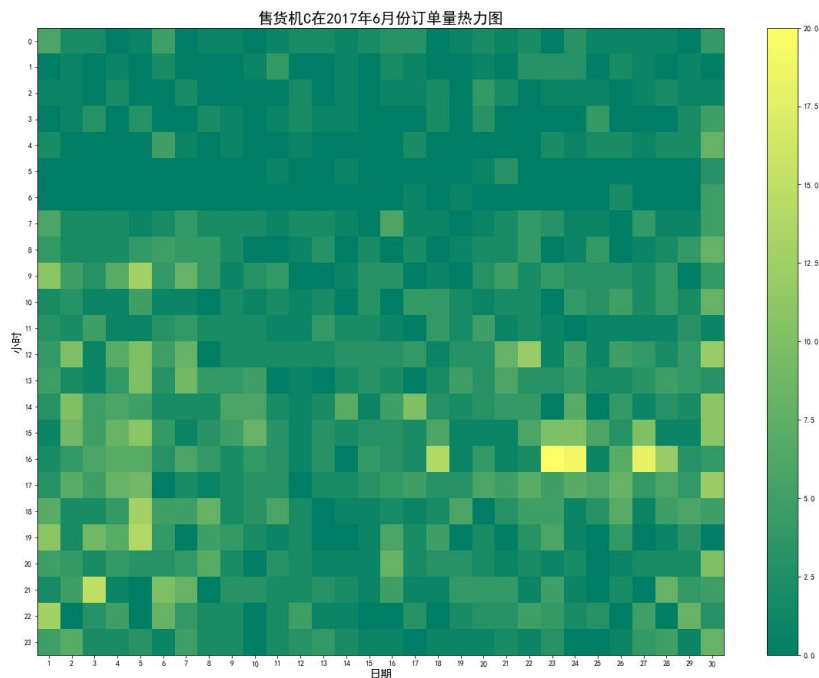


图 6 售货机 C 在 2017 年 6 月份订单量的热力图

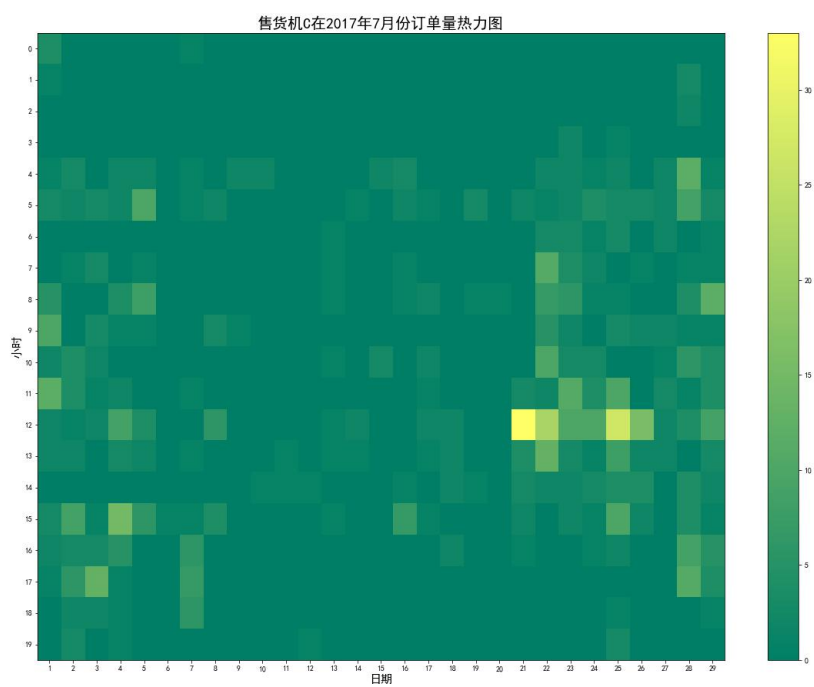


图 7 售货机 C 在 2017 年 7 月份订单量的热力图

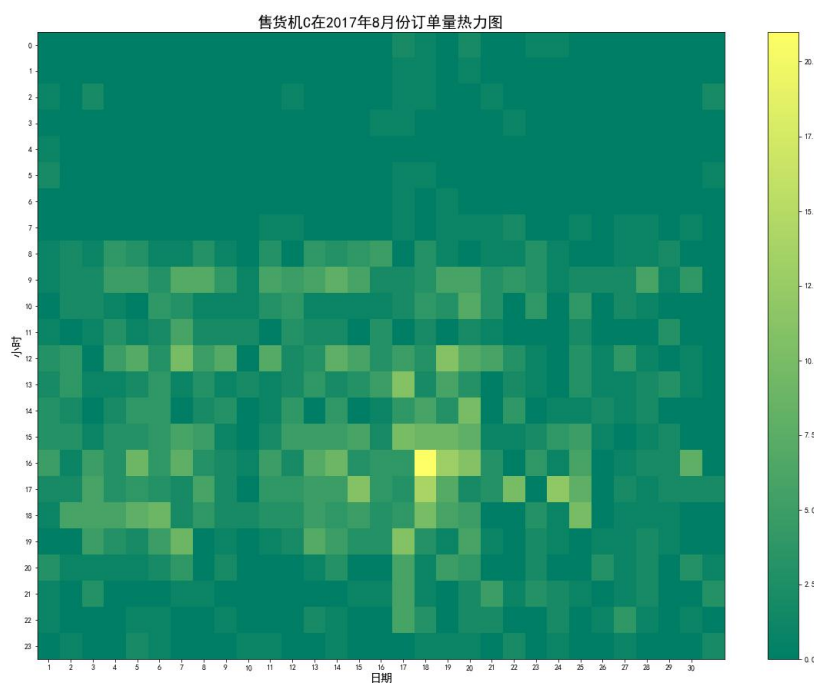


图 8 售货机 C 在 2017 年 8 月份订单量的热力图

图 6 至图 8 分别为售货机 C 在 2017 年 6、7、8 月份订单量的热

力图，横轴以天为单位，纵轴以小时为单位，图例表示颜色越黄订单量越多，颜色越绿订单量越少。

由图 6、图 7 颜色深浅比较得知，售货机 C 在 2017 年 6 月的订单量比 7 月的订单量多，其中 6 月的订单量分布比较复杂，6 月 23 和 24 日下午 4 点这一小时订单量最多，并且大多数是在早上 9 点至晚上 10 点出现订单量，其他时间段都有较少的订单量出现；售货机 C 在 2017 年 7 月份订单量分布较 6 月份明显稀疏了很多，7 月 21 号中午 12 点订单量这一小时订单量最多，其达到 7 月份的最大值，7 月 22 号至 31 号出现订单量数量较多，也是大多数在早上 9 点至晚上 10 点出现订单量，其他时间段都有较少的订单量出现。

由图 6 与图 8 比较可知，售货机 C 在 2017 年 8 月份的订单量与 6 月份的差不多，并且分布比较均匀的，每天的订单量不差上下，并且订单量主要集中在早上 8 点至晚上 10 点，在 8 月 18 号下午 4 点这一小时内订单量达到 8 月份的最大值。

由图 6 至图 8 可发现，订单分布通常在早上 8、9 点至晚上 10 点，由此可以推测出商场从早上 8 点开始聚集人流，直至晚上 10 点开始人流量减少。

## 五、 自动售货机画像

### 1. 贴标签

为了分析各售货机商品销售数据，总结规律，给每台售货机饮料类商品贴上标签：滞销/正常/热销，定义了“销量占比”的指标：

$$\text{销量占比} = \frac{\text{该饮料商品的销量}}{\text{所有饮料类商品的销量}}$$

算出所有饮料类商品的销量占比，然后按销售占比从大到小排序，使用 `quantile()` 函数分别计算出该序列的 20% 分位数和 80% 分位数，再定义销量占比小于 20% 分位数的商品标签为“滞销”，销量占比大于 80% 分位数的商品标签为“畅销”，销量占比介于 20% 分位数和 80% 分位数之间的商品标签是“正常”，因此每台售货机饮料类商品贴上了属于自己的标签，最后将对各售货机的所有饮料类商品的热销程度贴上标签保存在 csv 文件中，文件名分别为

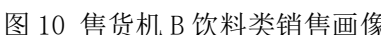
“task3-1A.csv”、“task3-1B.csv”、“task3-1C.csv”、“task3-1D.csv”、“task3-1E.csv”。

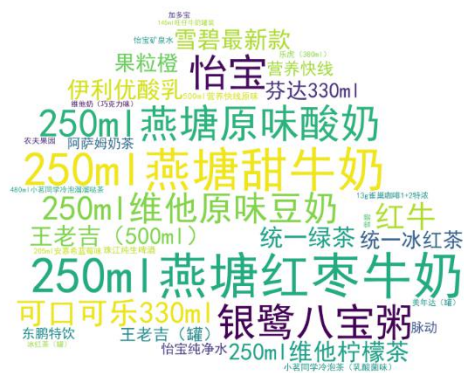
## 2. 生成各自动售货机画像

先导入所需 `wordcloud` 模块，并为各台售货机选择特定的背景图，然后依据商品的标签绘制各自动售货机的饮料类销售词云图，词云图是对文本中出现频率较高的关键词予以视觉化的展现，词云图过滤掉大量的低频低质的文本信息，使得浏览者只要一眼扫过文本就可以通过关键词来领略文本的主旨，如下是各台售货机的饮料类销售词云图：











根据词云图字体的大小，可以直观地看到：售货机 A 的热销饮料有雪碧最新款、伊利纯牛奶等；售货机 B 的热销饮料有银鹭八宝粥、雪碧最新款、伊利纯牛奶、可口可乐 330ml、145ml 旺子牛奶盒装、250ml 燕塘原味酸奶等；售货机 C 的热销饮料有 250ml 燕塘甜牛奶、

250ml 燕塘红枣牛奶、银鹭八宝粥、怡宝、250ml 燕塘原味酸奶等；售货机 D 的热销饮料有美年达(罐)、银鹭八宝粥等；售货机 E 的热销饮料有统一绿茶、伊利纯牛奶、145ml 旺子牛奶盒装、250ml 燕塘原味酸奶、脉动、可口可 330ml、伊利优酸乳、王老吉（罐）、东鹏特饮、红牛、银鹭八宝粥、雪碧最新款等。

可根据上述分析的热销饮料情况，可以对自动售货机内商品的频率、种类选择供给量进行选择，借此帮助经营者了解用户需求，掌握商品需求量，为用户提供精准贴心的服务，比如售货机 A、B 可以多存放雪碧最新款、伊利纯牛奶，售货机 C、D、E 可以多存放银鹭八宝粥等热销饮料。

## 六、 销售额预测

### 1. 预测未来销售额的原理

#### 1.1 销售预测的定义

销售额预测是销售预测的一个重要组成部分，销售预测是在充分考虑未来各种影响因素的基础上，根据历史销售资料以及市场上对产品需求的变化情况，对未来一定时期内有关产品的销售发展变化趋势所进行的科学预计和推测。

#### 1.2 销售额预测的影响因素

影响销售额预测的因素有很多，结合一般规律，本项目新零售无人智能售货机商务数据分析之预测未来销售额受以下因素影响：

（1）需求动向。需求是所有因素之中最重要的一项，如商场的人流量、

商场的活动举办、生活变化等，均可成为商品需求的影响因素，因此，必须加以分析与预测。

（2）价格变动。销售额深受价格变动的影 响，价格因素是影响商品销售的重要因素，为了提高销售额预测的准确性，应特别关注商品在自动售货机中的价格变动，以了解商品在自动售货机上的供应和需求情况。

（3）营销策略。不同的市场定位、价格政策、促销与服务政策等都会产生不同的销售额，因此要关注市场定位、价格政策、促销与服务政策等变更对销售额所产生的影响。

（4）同行竞争。销售额的高低深受同业竞争者的影响，预测自动售货机销售额需要了解竞争对手，掌握竞争对手的缺点与自身的优势，了解客户的真正需求，保持自身优势。

### 1.3 销售额预测的定量方法

销售额预测的方法主要有定性预测法和定量预测法两大类，但是本项目适用于定量预测法，因此此处只谈定量预测法。

用来进行销售预测的定量预测方法可以按照不同类型分成两大类：（1）时间序列分析法。时间序列分析法是利用变量与时间存在的相关关系，通过对以前数据的分析来预测将来的数据，在分析销售额时，将销售额按照年或月的次序排列下来，以观察其变化趋势；

（2）回归分析法。各种事物彼此之间都存在直接或间接的因果关系，同样的，销售量亦会随着某种变量的变化而变化。当销售与时间之外的其他事物存在相关性时，就可运用回归和相关分析法进行销售预

测。

## 1.4 销售额预测的模型

### a. 趋势外推预测方法

趋势外推预测方法是根据事物的历史和现实数据，寻求事物随时间推移而发展变化的规律，从而推测其未来状况的一种常用的预测方法。

趋势外推法的假设条件是：(1)假设事物发展过程没有跳跃式变化，即事物的发展变化是渐进型的；(2)假设所研究系统的结构、功能等基本保持不变，即假定根据过去资料建立的趋势外推模型能适合未来，能代表未来趋势变化的情况。由以上两个假设条件可知，趋势外推预测法是事物发展渐进过程的一种统计预测方法。简言之，就是运用一个数学模型，拟合一条趋势线，然后用这个模型外推预测未来时期事物的发展。

趋势外推预测法主要利用描绘散点图的方法(图形识别)和差分法计算进行模型选择。其主要优点是：可以揭示事物发展的未来，并定量地估价其功能特性。

趋势外推预测法比较适合中、长期新产品预测，要求有至少 5 年的数据资料。

### b. 回归预测方法

回归预测方法是根据自变量和因变量之间的相关关系进行预测的。自变量的个数可以一个或多个，根据自变量的个数可分为一元回

归预测和多元回归预测，同时根据自变量和因变量的相关关系，分为线性回归预测方法和非线性回归方法。

回归问题的学习等价于函数拟合：选择一条函数曲线使其很好的拟合已知数据且能很好的预测未知数据。

#### c. 卡尔曼滤波预测模型

卡尔曼滤波是以最小均方误差为估计的最佳准则，来寻求一套递推估计的模型，其基本思想是：采用信号与噪声的状态空间模型，利用前一时刻地估计值和现时刻的观测值来更新对状态变量的估计，求出出现时刻的估计值。

#### d. 组合预测模型

组合预测法是对同一个问题，采用多种预测方法，组合的主要目的是综合利用各种方法所提供的信息，尽可能地提高预测精度。组合预测有 2 种基本形式，一是等权组合，即各预测方法的预测值按相同的权数组合成新的预测值；二是不等权组合，即赋予不同预测方法的预测值不同的权数。这 2 种形式的原理和运用方法完全相同，只是权数的取定有所区别。根据经验，采用不等权组合的组合预测法结果较为准确。

#### e. BP 神经网络预测模型

BP 网络又称反向传播神经网络，通过样本数据的训练，不断修正网络权值和阈值使误差函数沿负梯度方向下降，逼近期望输出。它是一种应用较为广泛的神经网络模型，多用于函数逼近、模型识别分类、数据压缩和时间序列预测等。

## 2. 关于本案例中销售额的预测

本案例中的数据不足以对每台售货机的每大类商品在 2018 年 1 月的交易额进行预测，具体分析如下：

时间序列是按时间顺序的一组数字序列，本案例的数据符合时间序列分析法的基本特征，根据过去的变化趋势预测未来的发展。

时间序列分析的基本原理：一是承认事物发展的延续性，应用过去数据，就能推测事物的发展趋势。二是考虑到事物发展的随机性，任何事物发展都可能受偶然因素影响，为此要利用统计分析中加权平均法对历史数据进行处理，该方法简单易行，便于掌握，但准确性差，一般只适用于短期预测。本案例要求预测 2018 年 1 月各自动售货机的交易额，属于短期期预。

该时间序列中的每个观察值即各台自动售货机每月的销售额，是各种不同因素在同一时刻发生作用的综合结果。由数据可视化可以发现，该时间序列的要素有以下特点：

- a) 季节变动：2017 年自动售货机的的销售额由于自然季节的天气等其他原因交替出现高峰与低谷的规律。
- b) 循环波动：2017 年各自动售货机的销售额涨落相同。
- c) 趋势：2017 年各自动售货机的销售额随着时间进展，整体呈现上升趋势。
- d) 不规则波动：个别偶然因素对时间序列产生影响，没有规律。

因此，根据传统时间序列分析法原理上可以预测自动售货机未来销售额，但是计算时间序列的成分需要连续 2 年或者 2 年以上的有关



销售数据，而本案例只提供了 1 年的数据，即使预测出来，准确性也是非常令人担忧的，因此无法对每台售货机的每大类商品在 2018 年 1 月的交易额进行预测。

如果要预测自动售货机未来销售额，经营者需要提供连续 2 年或者 2 年以上的有关自动售货机销售数据，而且要保证数据的针对性、真实性、完整性以及可比性，这对预测结果的准确度具有重要的影响，还要有关自动售货机的市场需求潜量、市场占有率、产品的售价等统计资料，以掌握市场的需求动向等等。

## 七、 项目总结

本次项目是对新零售无人智能售货机商务数据进行分析。对此，先对该数据进行预处理，解决脏数据，整理数据，计算所需数据，再将数据可视化，并用不同的图形表示，然后为每台售货机所销售的商品贴上标签，并绘制各台自动售货机的饮料类的销售画像，最后由于所提供的数据不足以对每台售货机的每个大类商品在 2018 年 1 月的交易额进行预测，所以对经营者提出了关于数据方面的要求。